Graphs

Bing Xu

11/29/2016

```r
require(tidyverse)
require(stringr)
require(forcats)
require(ggplot2)
require(choroplethr)
require(choroplethrMaps)
require(dplyr)
library(tidyverse)

dat<-read.csv("CC-EST2015-ALLDATA.csv",header = TRUE)

##FOR ALL AGE GROUPS AND FOR YEAR 2015
dat2015<-dat %>%
  filter(AGEGRP==0 & YEAR==8)

##DO NOT COUNT HISPANIC
##SIMPLY USING TWO 6 TYPES OF RACES AND THIS COULD BE REVISED
##H=H_MALE+H_FEMALE,
dat2015<-dat2015 %>%
  transmute(STATE=STATE,
            COUNTY=COUNTY,
            TOT=TOT_POP,
            WA=WA_MALE+WA_FEMALE,
            BA=BA_MALE+BA_FEMALE,
            IA=IA_MALE+IA_FEMALE,
            AA=AA_MALE+AA_FEMALE,
            NHA=NA_MALE+NA_FEMALE,
            TOM=TOM_MALE+TOM_FEMALE,
            WAP=WA/TOT,
            BAP=BA/TOT,
            IAP=IA/TOT,
            AAP=AA/TOT,
            NHAP=NHA/TOT,
            TOMP=TOM/TOT,
            INDEX=1-WAP^2-BAP^2-IAP^2-AAP^2-NHAP^2-TOMP^2)

dat2015<-dat2015 %>%
  mutate(STATE_char = as.character(STATE), COUNTY_char = as.character(COUNTY))
dat2015$STATE_char = str_pad(dat2015$STATE_char, 2, pad = "0")
dat2015$COUNTY_char = str_pad(dat2015$COUNTY_char, 3, pad = "0")

dat2015<-dat2015 %>%
  mutate(FIPS = paste(STATE_char, COUNTY_char, sep = ""))

##FOR YEAR 2010
dat2010<-dat %>%
  filter(AGEGRP==0 & YEAR==1)
```

```r
dat2010<-dat2010 %>%
  transmute(STATE=STATE,
            COUNTY=COUNTY,
            TOT=TOT_POP,
            WA=WA_MALE+WA_FEMALE,
            BA=BA_MALE+BA_FEMALE,
            IA=IA_MALE+IA_FEMALE,
            AA=AA_MALE+AA_FEMALE,
            NHA=NA_MALE+NA_FEMALE,
            TOM=TOM_MALE+TOM_FEMALE,
            WAP=WA/TOT,
            BAP=BA/TOT,
            IAP=IA/TOT,
            AAP=AA/TOT,
            NHAP=NHA/TOT,
            TOMP=TOM/TOT,
            INDEX=1-WAP^2-BAP^2-IAP^2-AAP^2-NHAP^2-TOMP^2)

dat2010<-dat2010 %>%
  mutate(STATE_char = as.character(STATE), COUNTY_char = as.character(COUNTY))
dat2010$STATE_char = str_pad(dat2010$STATE_char, 2, pad = "0")
dat2010$COUNTY_char = str_pad(dat2010$COUNTY_char, 3, pad = "0")
dat2010<-dat2010 %>%
  mutate(FIPS = paste(STATE_char, COUNTY_char, sep = ""))
dat2010$FIPS<-as.numeric(dat2010$FIPS)

##FOR YEAR 2011
dat2011<-dat %>%
  filter(AGEGRP==0 & YEAR==4)

dat2011<-dat2011 %>%
  transmute(STATE=STATE,
            COUNTY=COUNTY,
            TOT=TOT_POP,
            WA=WA_MALE+WA_FEMALE,
            BA=BA_MALE+BA_FEMALE,
            IA=IA_MALE+IA_FEMALE,
            AA=AA_MALE+AA_FEMALE,
            NHA=NA_MALE+NA_FEMALE,
            TOM=TOM_MALE+TOM_FEMALE,
            WAP=WA/TOT,
            BAP=BA/TOT,
            IAP=IA/TOT,
            AAP=AA/TOT,
            NHAP=NHA/TOT,
            TOMP=TOM/TOT,
            INDEX=1-WAP^2-BAP^2-IAP^2-AAP^2-NHAP^2-TOMP^2)

dat2011<-dat2011 %>%
  mutate(STATE_char = as.character(STATE), COUNTY_char = as.character(COUNTY))
dat2011$STATE_char = str_pad(dat2011$STATE_char, 2, pad = "0")
dat2011$COUNTY_char = str_pad(dat2011$COUNTY_char, 3, pad = "0")
dat2011<-dat2011 %>%
```

```r
  mutate(FIPS = paste(STATE_char, COUNTY_char, sep = ""))
dat2011$FIPS<-as.numeric(dat2011$FIPS)

##For Year 2012
dat2012<-dat %>%
  filter(AGEGRP==0 & YEAR==5)

dat2012<-dat2012 %>%
  transmute(STATE=STATE,
            COUNTY=COUNTY,
            TOT=TOT_POP,
            WA=WA_MALE+WA_FEMALE,
            BA=BA_MALE+BA_FEMALE,
            IA=IA_MALE+IA_FEMALE,
            AA=AA_MALE+AA_FEMALE,
            NHA=NA_MALE+NA_FEMALE,
            TOM=TOM_MALE+TOM_FEMALE,
            WAP=WA/TOT,
            BAP=BA/TOT,
            IAP=IA/TOT,
            AAP=AA/TOT,
            NHAP=NHA/TOT,
            TOMP=TOM/TOT,
            INDEX=1-WAP^2-BAP^2-IAP^2-AAP^2-NHAP^2-TOMP^2)

dat2012<-dat2012 %>%
  mutate(STATE_char = as.character(STATE), COUNTY_char = as.character(COUNTY))
dat2012$STATE_char = str_pad(dat2012$STATE_char, 2, pad = "0")
dat2012$COUNTY_char = str_pad(dat2012$COUNTY_char, 3, pad = "0")
dat2012<-dat2012 %>%
  mutate(FIPS = paste(STATE_char, COUNTY_char, sep = ""))
dat2012$FIPS<-as.numeric(dat2012$FIPS)

##For Year 2013
dat2013<-dat %>%
  filter(AGEGRP==0 & YEAR==6)

dat2013<-dat2013 %>%
  transmute(STATE=STATE,
            COUNTY=COUNTY,
            TOT=TOT_POP,
            WA=WA_MALE+WA_FEMALE,
            BA=BA_MALE+BA_FEMALE,
            IA=IA_MALE+IA_FEMALE,
            AA=AA_MALE+AA_FEMALE,
            NHA=NA_MALE+NA_FEMALE,
            TOM=TOM_MALE+TOM_FEMALE,
            WAP=WA/TOT,
            BAP=BA/TOT,
            IAP=IA/TOT,
            AAP=AA/TOT,
            NHAP=NHA/TOT,
            TOMP=TOM/TOT,
```

```
                INDEX=1-WAP^2-BAP^2-IAP^2-AAP^2-NHAP^2-TOMP^2)

dat2013<-dat2013 %>%
  mutate(STATE_char = as.character(STATE), COUNTY_char = as.character(COUNTY))
dat2013$STATE_char = str_pad(dat2013$STATE_char, 2, pad = "0")
dat2013$COUNTY_char = str_pad(dat2013$COUNTY_char, 3, pad = "0")
dat2013<-dat2013 %>%
  mutate(FIPS = paste(STATE_char, COUNTY_char, sep = ""))
dat2013$FIPS<-as.numeric(dat2013$FIPS)

##For Year 2014
dat2014<-dat %>%
  filter(AGEGRP==0 & YEAR==7)
dat2014<-dat2014 %>%
  transmute(STATE=STATE,
            COUNTY=COUNTY,
            TOT=TOT_POP,
            WA=WA_MALE+WA_FEMALE,
            BA=BA_MALE+BA_FEMALE,
            IA=IA_MALE+IA_FEMALE,
            AA=AA_MALE+AA_FEMALE,
            NHA=NA_MALE+NA_FEMALE,
            TOM=TOM_MALE+TOM_FEMALE,
            WAP=WA/TOT,
            BAP=BA/TOT,
            IAP=IA/TOT,
            AAP=AA/TOT,
            NHAP=NHA/TOT,
            TOMP=TOM/TOT,
            INDEX=1-WAP^2-BAP^2-IAP^2-AAP^2-NHAP^2-TOMP^2)

dat2014<-dat2014 %>%
  mutate(STATE_char = as.character(STATE), COUNTY_char = as.character(COUNTY))
dat2014$STATE_char = str_pad(dat2014$STATE_char, 2, pad = "0")
dat2014$COUNTY_char = str_pad(dat2014$COUNTY_char, 3, pad = "0")
head(dat2014)
```

```
##   STATE COUNTY    TOT     WA    BA   IA   AA NHA  TOM       WAP        BAP
## 1     1      1  55290  43034 10317  270  628  55  986 0.7783324 0.18659794
## 2     1      3 199713 173960 19143 1479 1767 120 3244 0.8710500 0.09585255
## 3     1      5  26815  13461 12781  154  125  49  245 0.5019952 0.47663621
## 4     1      7  22549  17206  4972   92   45  29  205 0.7630494 0.22049758
## 5     1      9  57658  55307  1032  365  166  64  724 0.9592251 0.01789864
## 6     1     11  10829   2919  7603   86   27  74  120 0.2695540 0.70209622
##          IAP         AAP         NHAP        TOMP     INDEX STATE_char
## 1 0.004883342 0.011358293 0.0009947549 0.017833243 0.3589080         01
## 2 0.007405627 0.008847696 0.0006008622 0.016243309 0.2316869         01
## 3 0.005743054 0.004661570 0.0018273354 0.009136677 0.5206773         01
## 4 0.004080004 0.001995654 0.0012860881 0.009091312 0.3690316         01
## 5 0.006330431 0.002879045 0.0011099934 0.012556800 0.0793596         01
## 6 0.007941638 0.002493305 0.0068335026 0.011081356 0.4341628         01
##   COUNTY_char
## 1         001
```

4

```
## 2               003
## 3               005
## 4               007
## 5               009
## 6               011
```

```
dat2014<-dat2014 %>%
  mutate(FIPS = paste(STATE_char, COUNTY_char, sep = ""))
head(dat2014)
```

```
##   STATE COUNTY    TOT     WA    BA   IA   AA NHA  TOM        WAP        BAP
## 1     1      1  55290  43034 10317  270  628  55  986 0.7783324 0.18659794
## 2     1      3 199713 173960 19143 1479 1767 120 3244 0.8710500 0.09585255
## 3     1      5  26815  13461 12781  154  125  49  245 0.5019952 0.47663621
## 4     1      7  22549  17206  4972   92   45  29  205 0.7630494 0.22049758
## 5     1      9  57658  55307  1032  365  166  64  724 0.9592251 0.01789864
## 6     1     11  10829   2919  7603   86   27  74  120 0.2695540 0.70209622
##           IAP         AAP         NHAP        TOMP     INDEX STATE_char
## 1 0.004883342 0.011358293 0.0009947549 0.017833243 0.3589080         01
## 2 0.007405627 0.008847696 0.0006008622 0.016243309 0.2316869         01
## 3 0.005743054 0.004661570 0.0018273354 0.009136677 0.5206773         01
## 4 0.004080004 0.001995654 0.0012860881 0.009091312 0.3690316         01
## 5 0.006330431 0.002879045 0.0011099934 0.012556800 0.0793596         01
## 6 0.007941638 0.002493305 0.0068335026 0.011081356 0.4341628         01
##   COUNTY_char  FIPS
## 1         001 01001
## 2         003 01003
## 3         005 01005
## 4         007 01007
## 5         009 01009
## 6         011 01011
```

```
dat2014$FIPS<-as.numeric(dat2014$FIPS)


##Distribution of Diversity Index
dat2010 %>%
  ggplot(aes(x=INDEX))+geom_histogram()
```
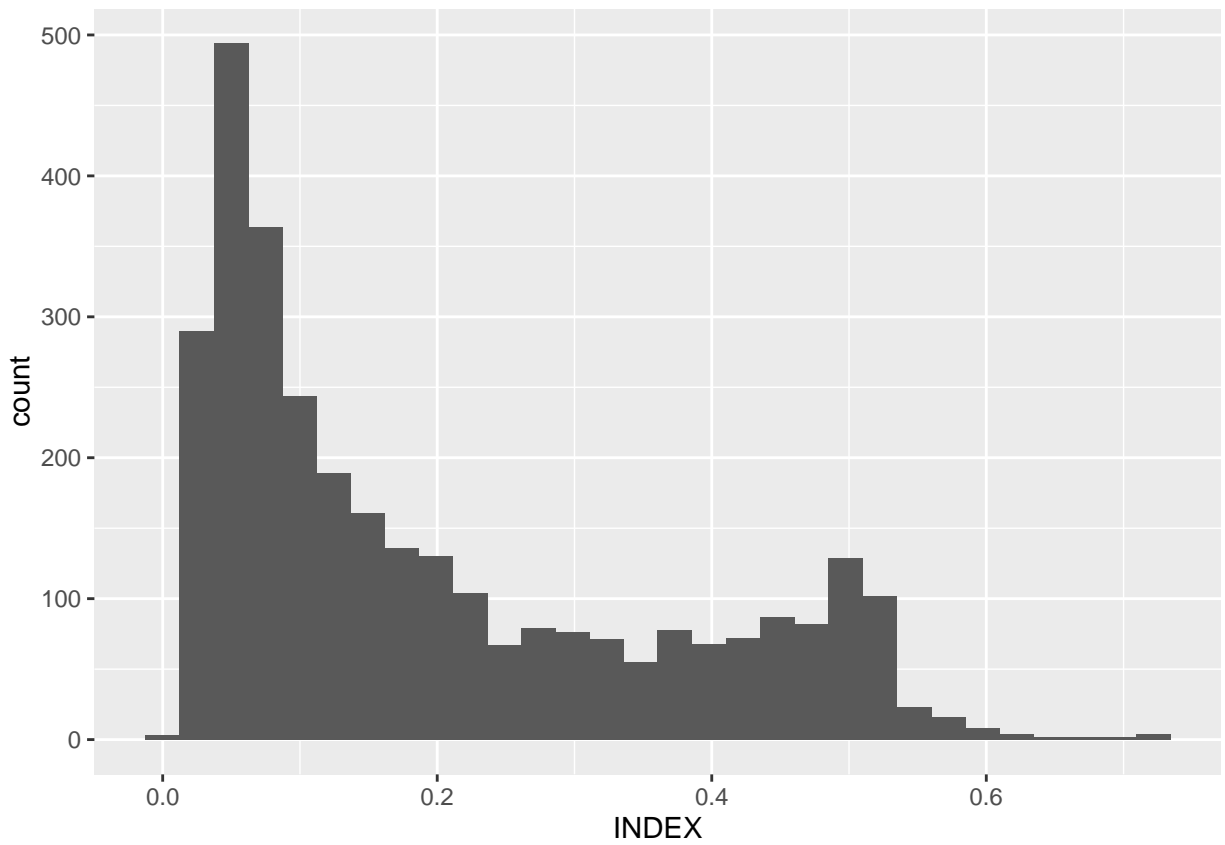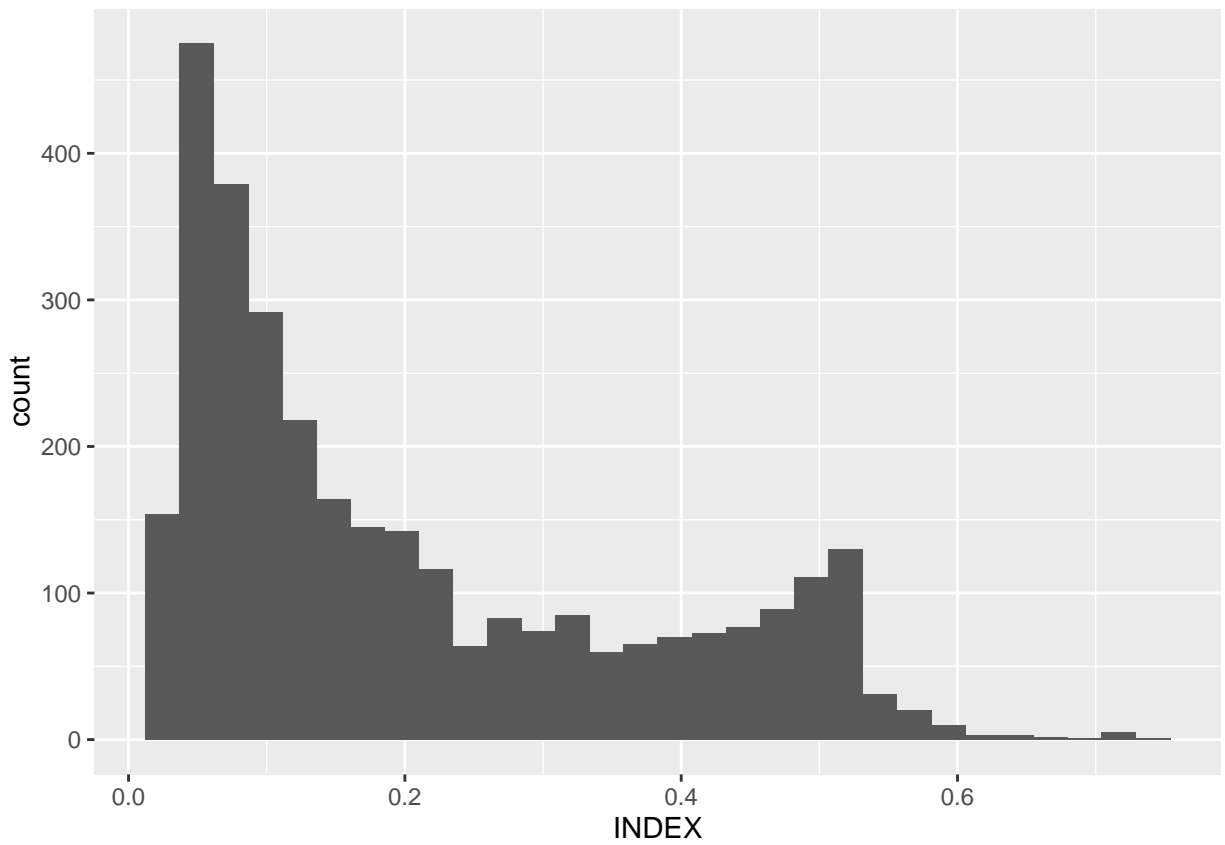
```
dat2011 %>%
  ggplot(aes(x=INDEX))+geom_histogram()
```
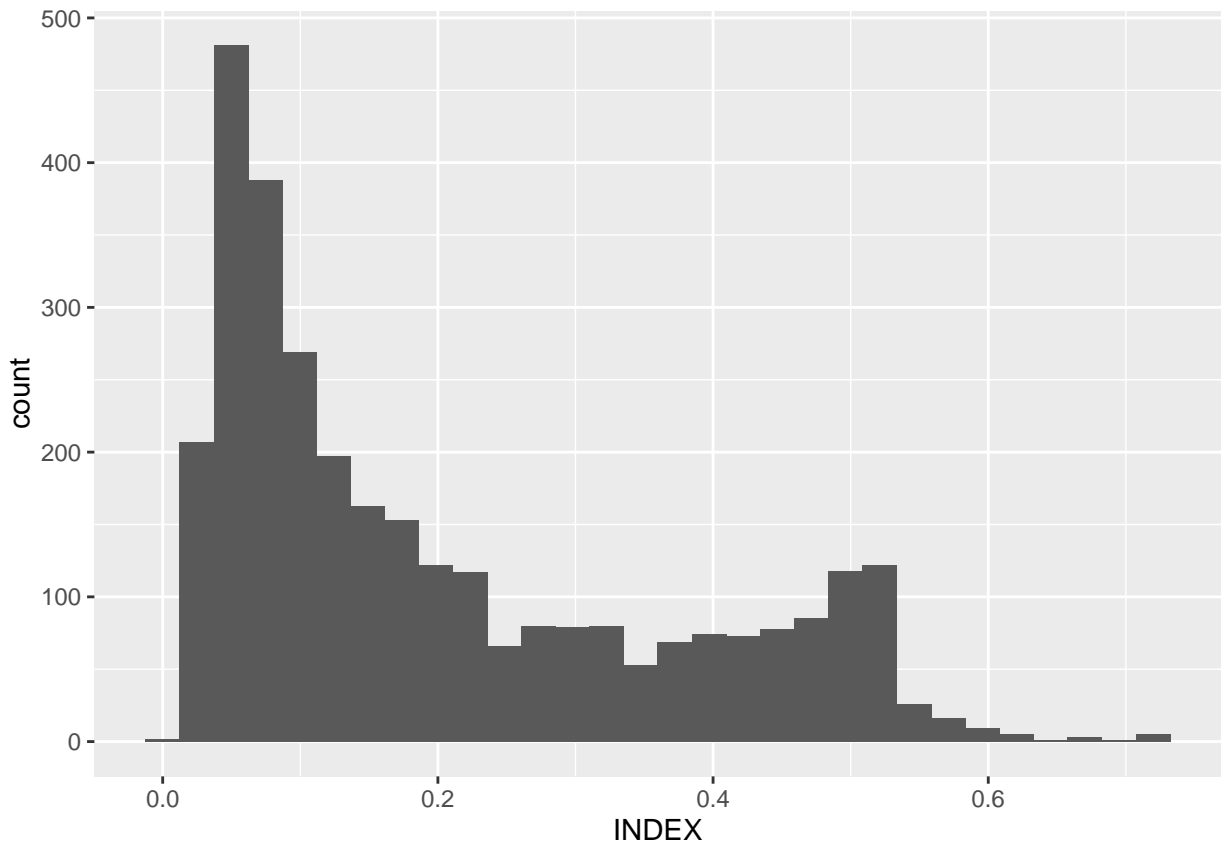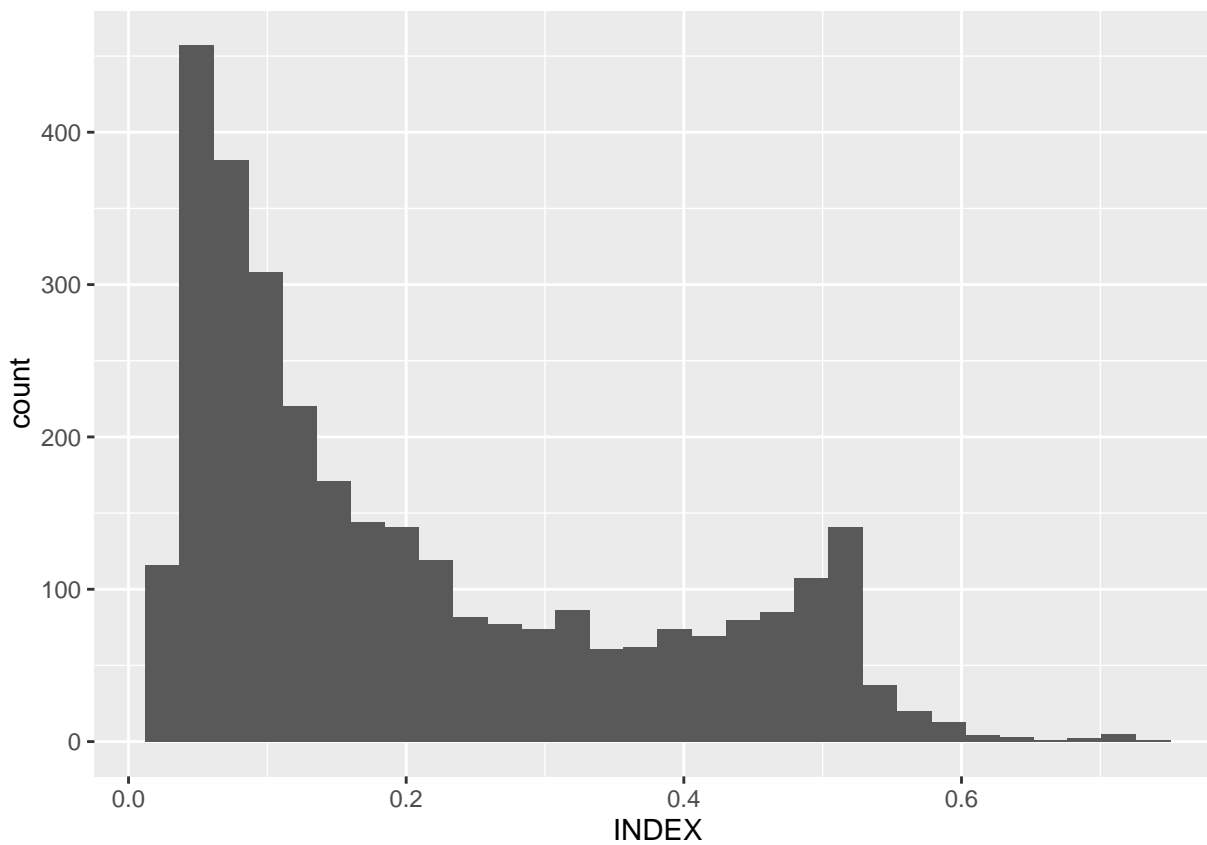
```
dat2012 %>%
  ggplot(aes(x=INDEX))+geom_histogram()
```
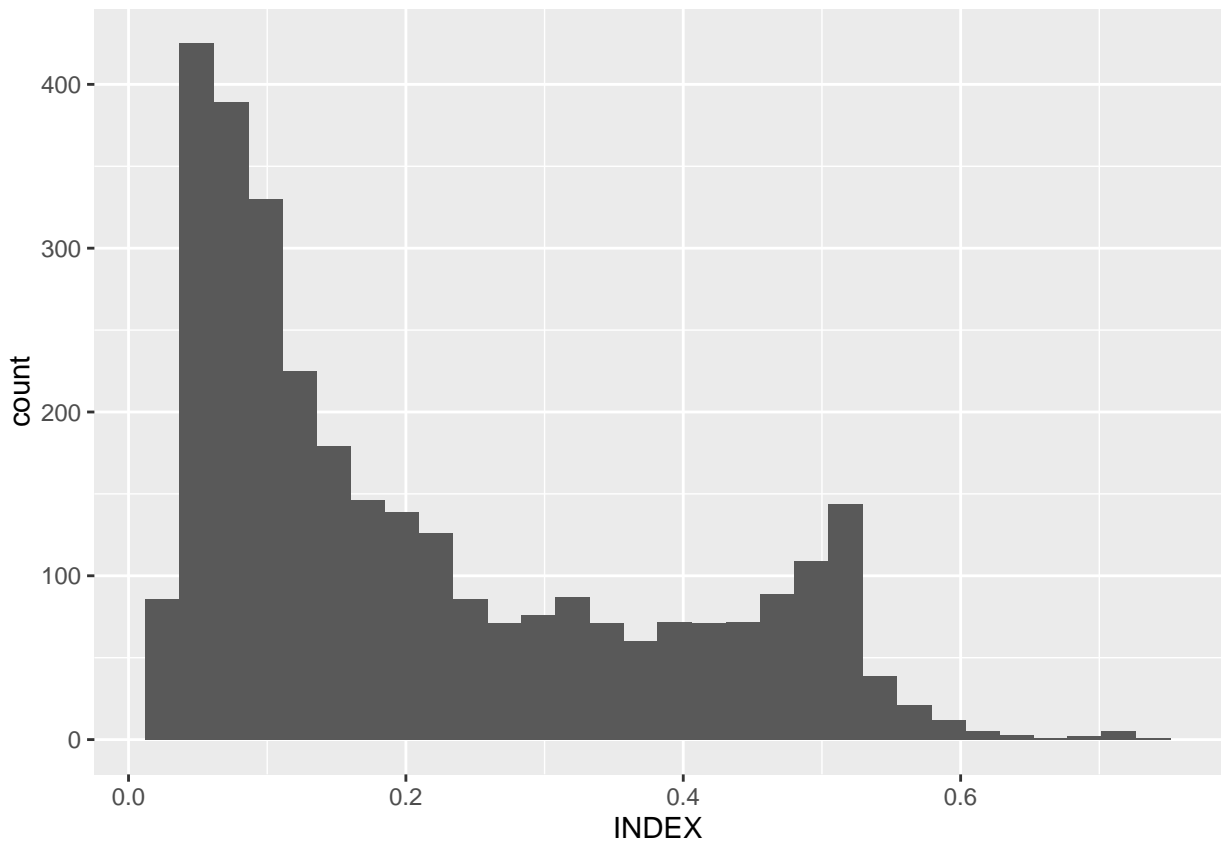
```
dat2013 %>%
  ggplot(aes(x=INDEX))+geom_histogram()
```
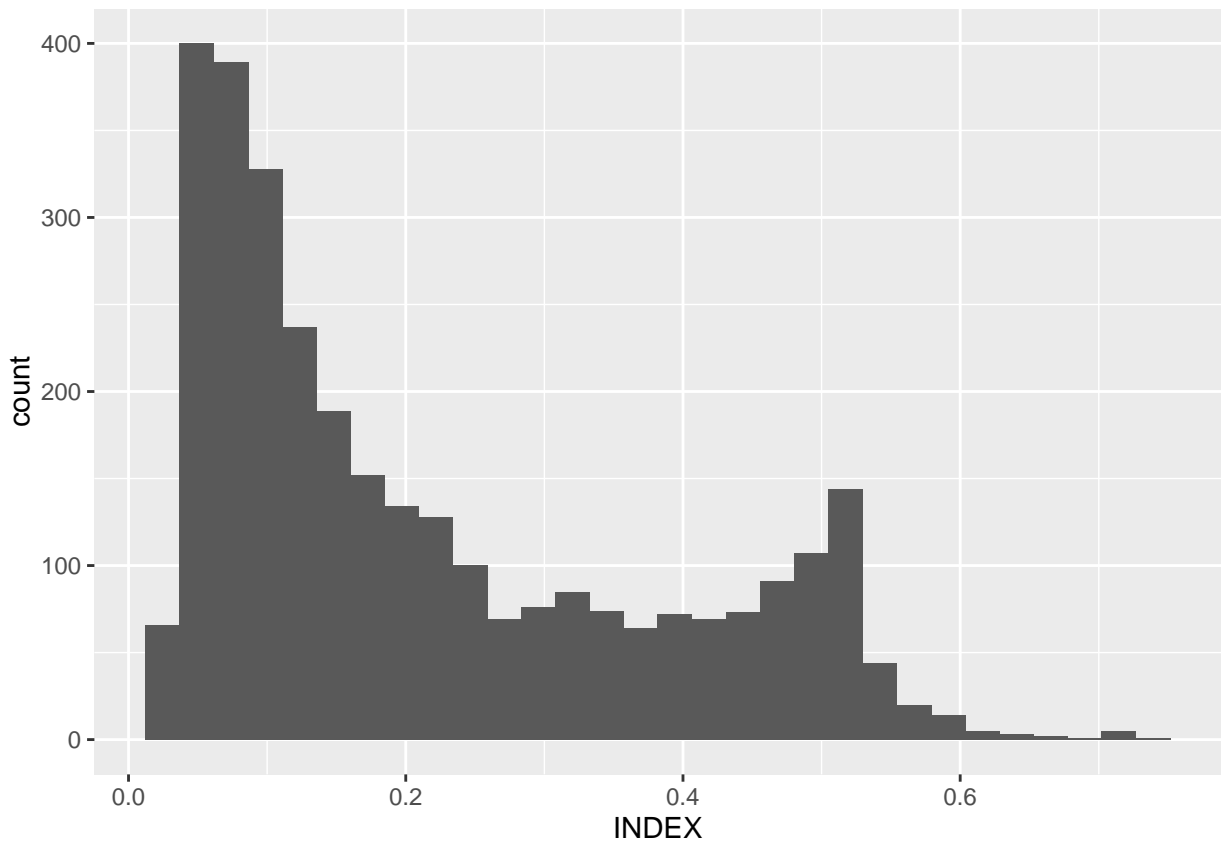
```
dat2014 %>%
  ggplot(aes(x=INDEX))+geom_histogram()
```
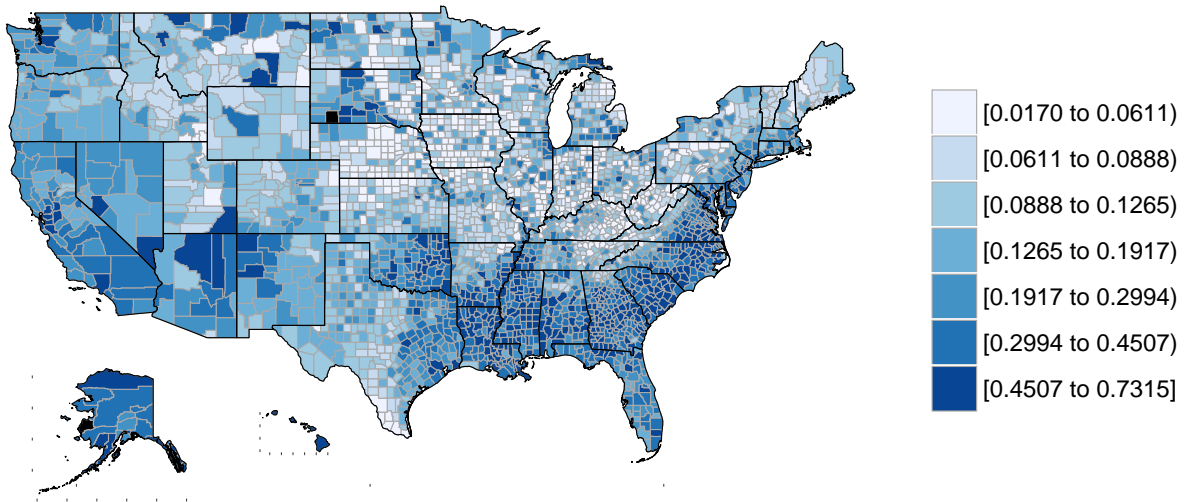
```
dat2015 %>%
  ggplot(aes(x=INDEX))+geom_histogram()
```

```
##Diversity Index Map
mapping <- select(dat2015, FIPS, INDEX)
mapping <- dplyr::rename(mapping, region=FIPS, value=INDEX)
mapping$region <- as.numeric(mapping$region)

county_choropleth(mapping, title = "Diversity Index in Counties Across the US")
```

Diversity Index in Counties Across the US

|	
	[0.0170 to 0.0611)
	[0.0611 to 0.0888)
	[0.0888 to 0.1265)
	[0.1265 to 0.1917)
	[0.1917 to 0.2994)
	[0.2994 to 0.4507)
	[0.4507 to 0.7315]

```
dat2015$FIPS<-as.numeric(dat2015$FIPS)
##Education using 2010 data
read.csv("Education.csv", header = TRUE, stringsAsFactors = FALSE) %>%
  tbl_df() ->
  edu

edu_comb<-left_join(dat2010, edu, by=c("FIPS"="FIPS.Code"))

edu_comb %>%
  ggplot(aes(x=INDEX,
             y=Percent.of.adults.with.less.than.a.high.school.diploma..2010.2014))+
  geom_point()+geom_smooth()+
  ylab("Percentage")+
  xlab("Diversity Index")+
  ggtitle("Diversity vs No High School Diploma Percentage")
```
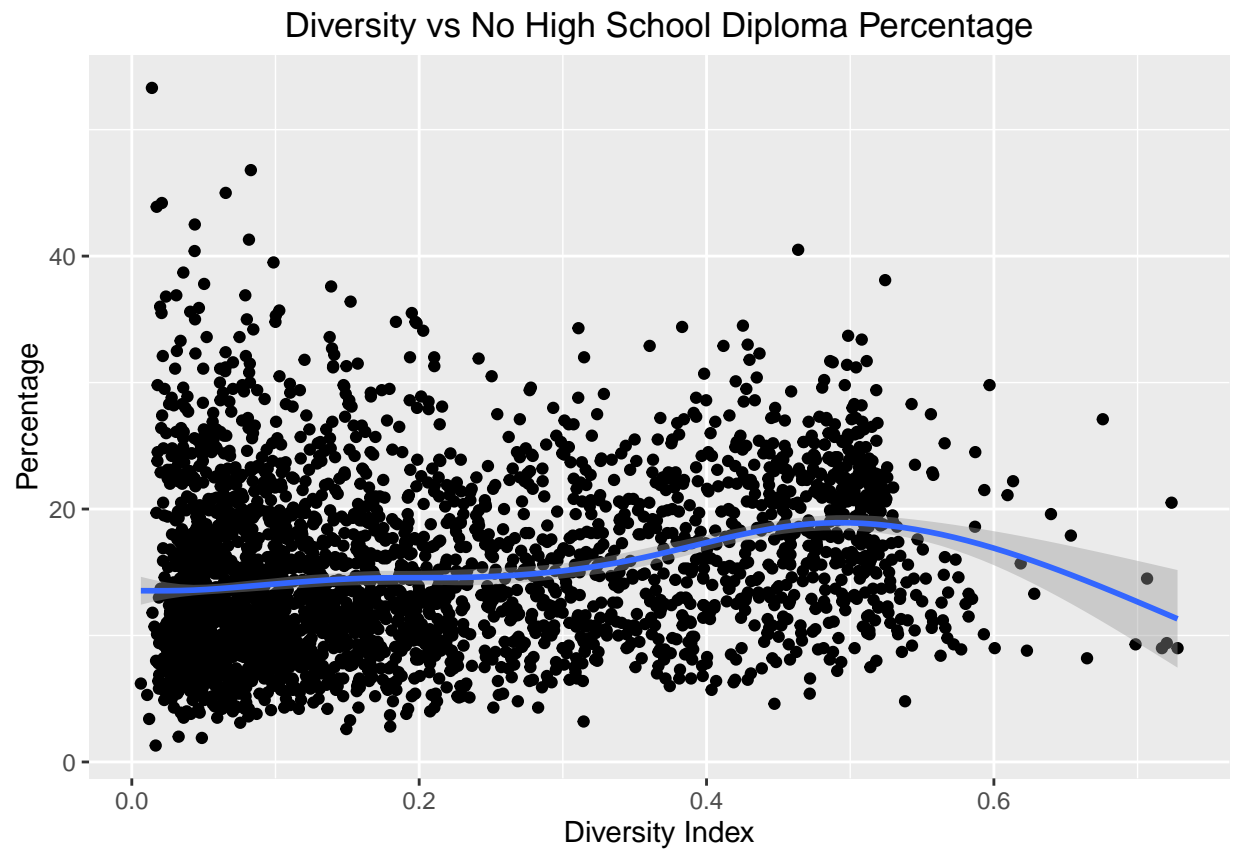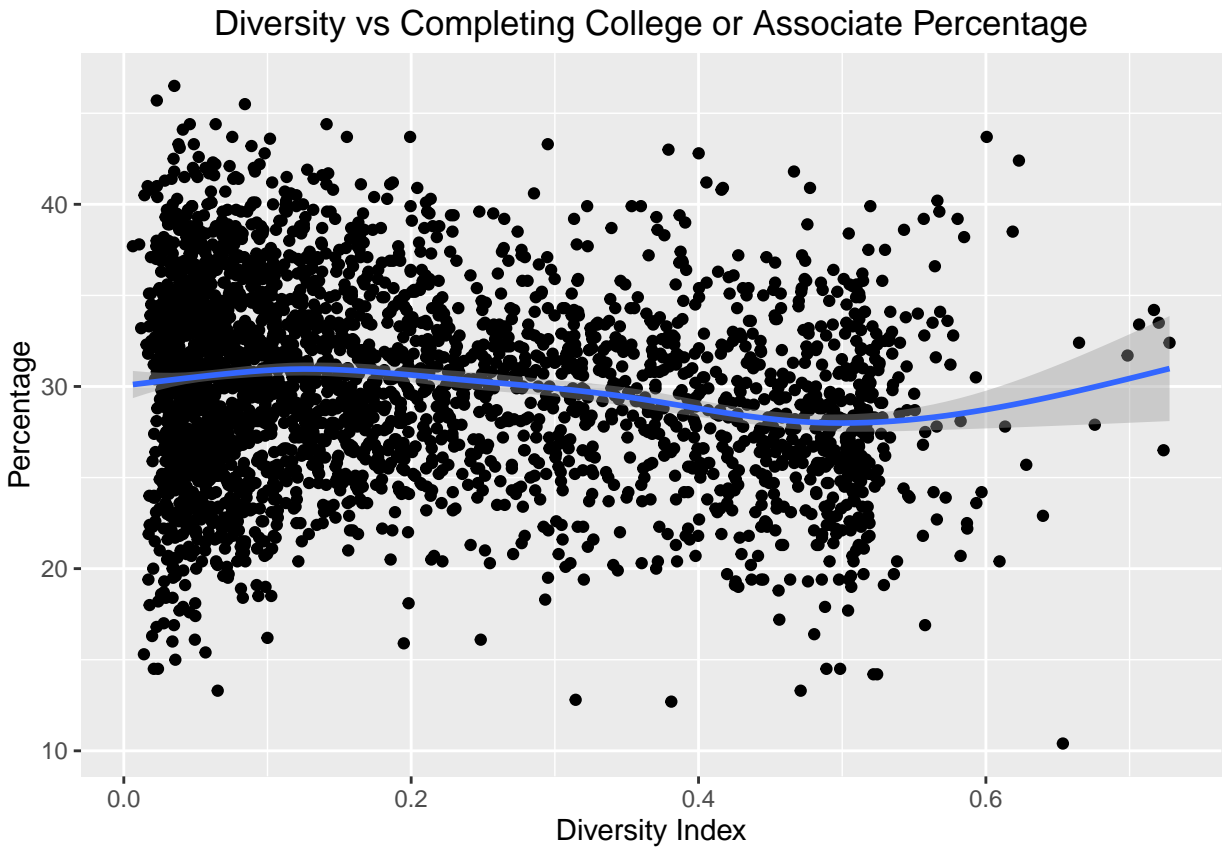
Diversity vs No High School Diploma Percentage


```
edu_comb %>%
  ggplot(aes(x=INDEX,
             y=Percent.of.adults.with.a.high.school.diploma.only..2010.2014))+
  geom_point()+geom_smooth()+
  ylab("Percentage")+
  xlab("Diversity Index")+
  ggtitle("Diversity vs High School Diploma Only Percentage")
```
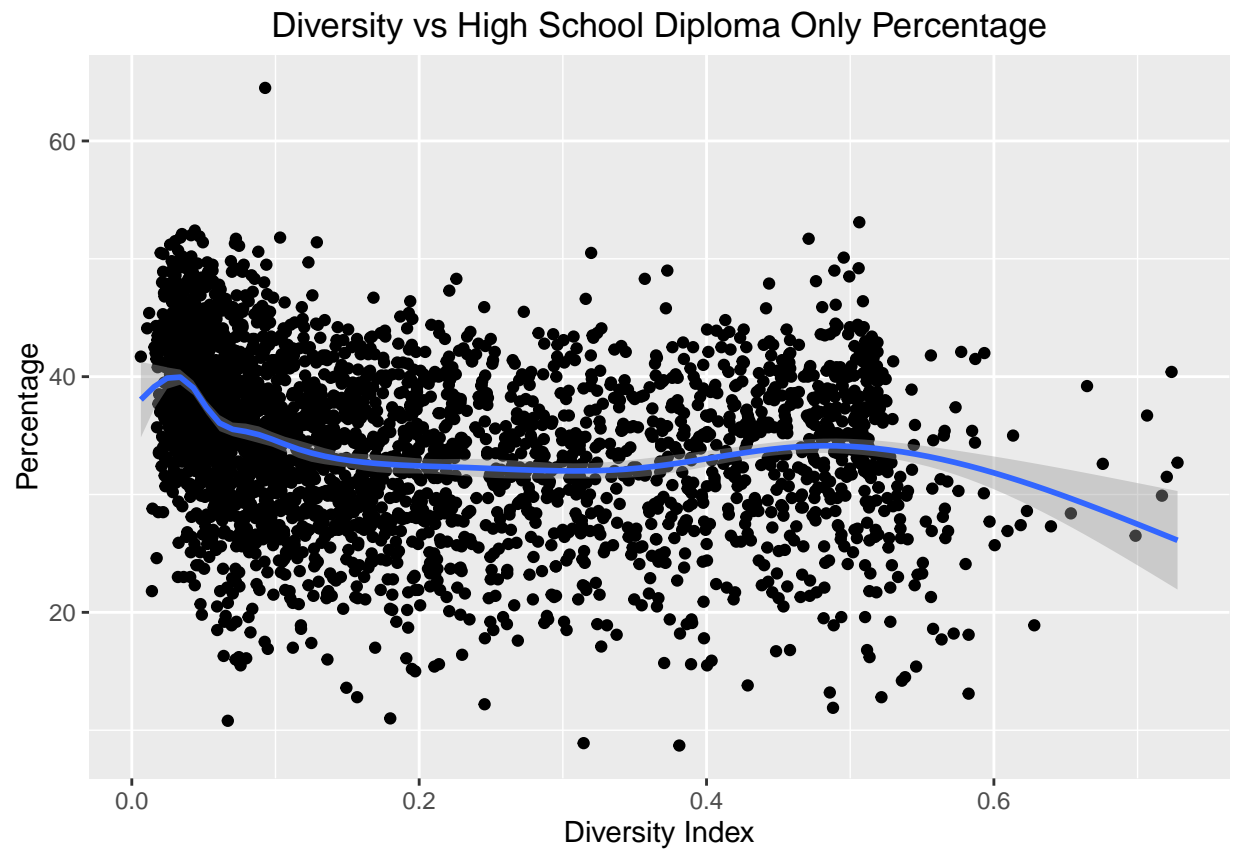
Diversity vs High School Diploma Only Percentage



```
edu_comb %>%
  ggplot(aes(x=INDEX,
             y=Percent.of.adults.completing.some.college.or.associate.s.degree..2010.2014))+
  geom_point()+geom_smooth()+
  ylab("Percentage")+
  xlab("Diversity Index")+
  ggtitle("Diversity vs Completing College or Associate Percentage")
```

Diversity vs Completing College or Associate Percentage


```
edu_comb %>%
  ggplot(aes(x=INDEX,
             y=Percent.of.adults.with.a.bachelor.s.degree.or.higher..2010.2014))+
  geom_point()+geom_smooth()+
  ylab("Percentage")+
  xlab("Diversity Index")+
  ggtitle("Diversity vs Bachelor Degree or Higher Percentage")
```

Diversity vs Bachelor Degree or Higher Percentage
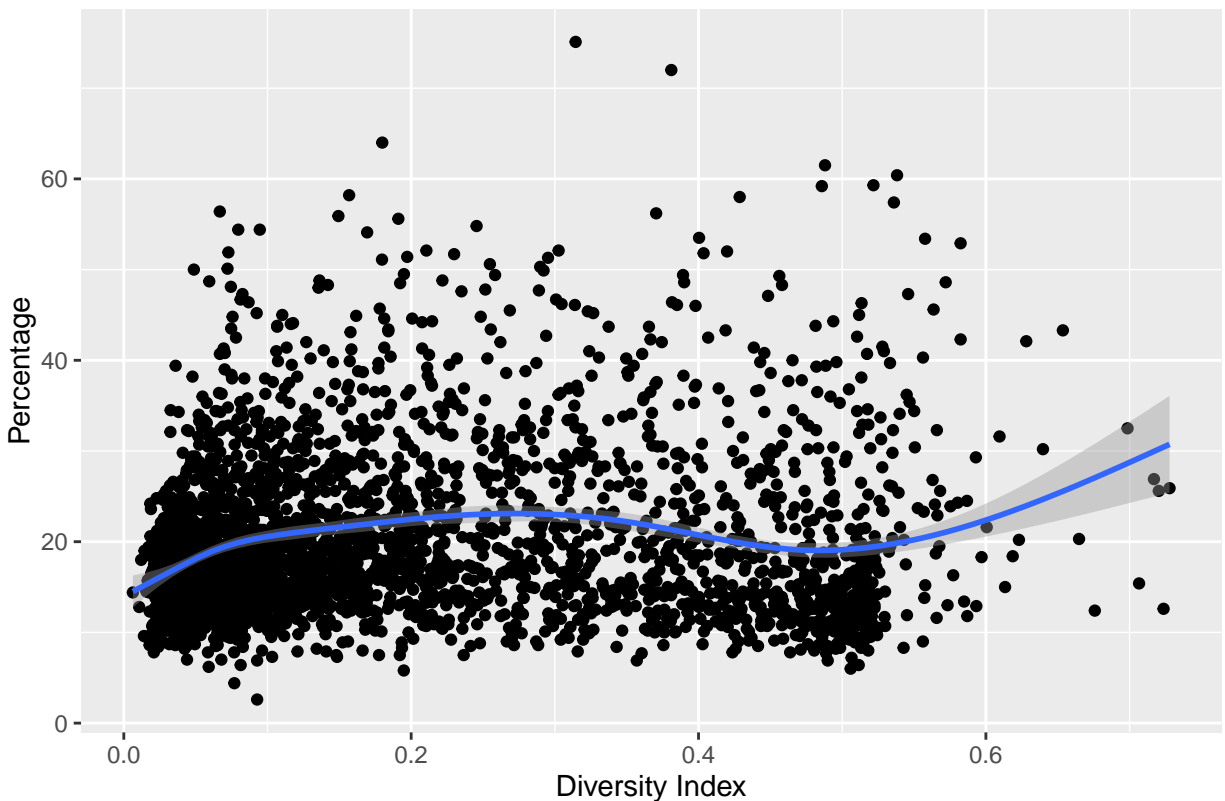

```
##Linear Regression Analysis
l1<-lm(Percent.of.adults.with.less.than.a.high.school.diploma..2010.2014~INDEX, data = edu_comb)
##summary(l1)
l2<-lm(Percent.of.adults.with.a.high.school.diploma.only..2010.2014~INDEX, data = edu_comb)
##summary(l2)
l3<-lm(Percent.of.adults.completing.some.college.or.associate.s.degree..2010.2014~INDEX, data = edu_com
##summary(l3)
l4<-lm(Percent.of.adults.with.a.bachelor.s.degree.or.higher..2010.2014~INDEX, data = edu_comb)
##summary(l4)

##Health
life<-read.csv("IHME_county_data_LifeExpectancy.csv", header = TRUE,stringsAsFactors = FALSE)

##Economic for year 2014
eco<-read.csv("Unemployment_and_Income1.csv",header = TRUE, stringsAsFactors = FALSE)
eco$Median_Household_Income_2014<-as.numeric(eco$Median_Household_Income_2014)
eco_comb2014<-left_join(dat2014, eco, by=c("FIPS"="FIPS_Code"))

##Income
eco_comb2014 %>%
  ggplot(aes(x=INDEX,
             y=Median_Household_Income_2014))+
  geom_point()+geom_smooth()+
  ylab("Median Income")+
  xlab("Diversity Index")+
  ggtitle("The Relationship between Diversity and Income in 2014")
```
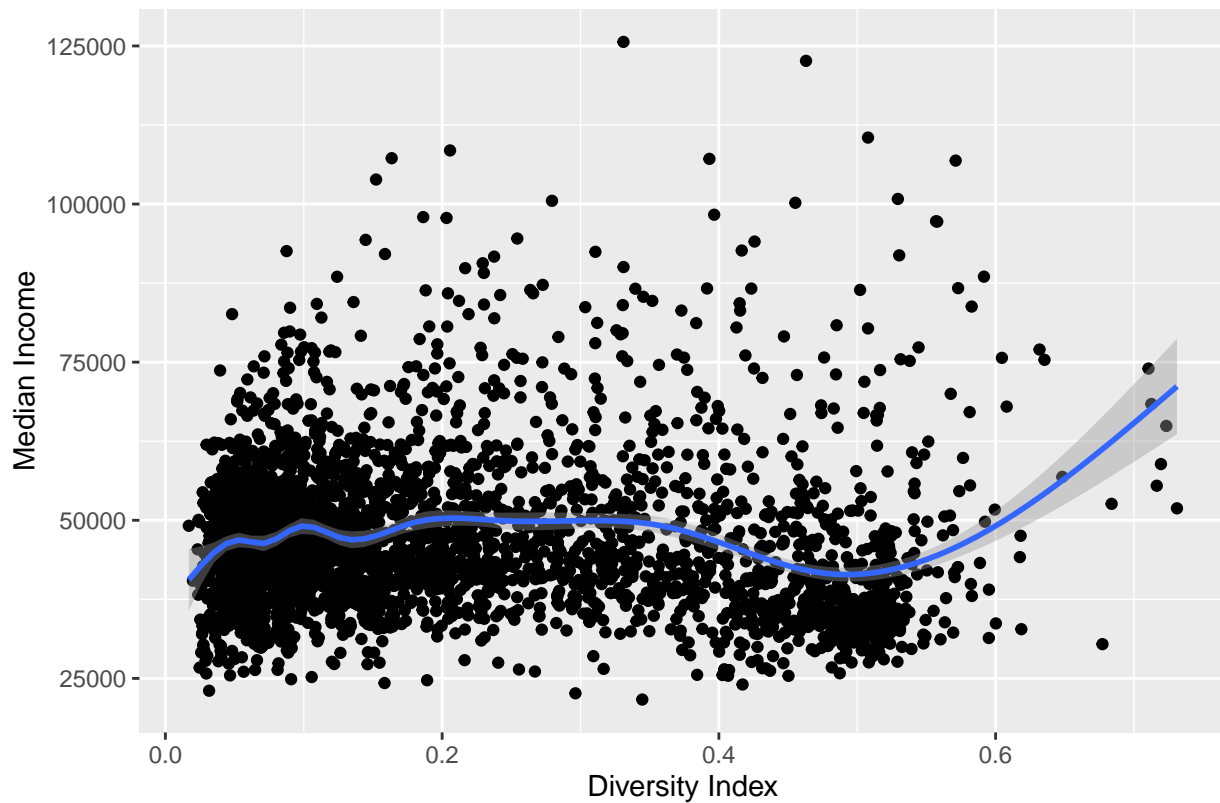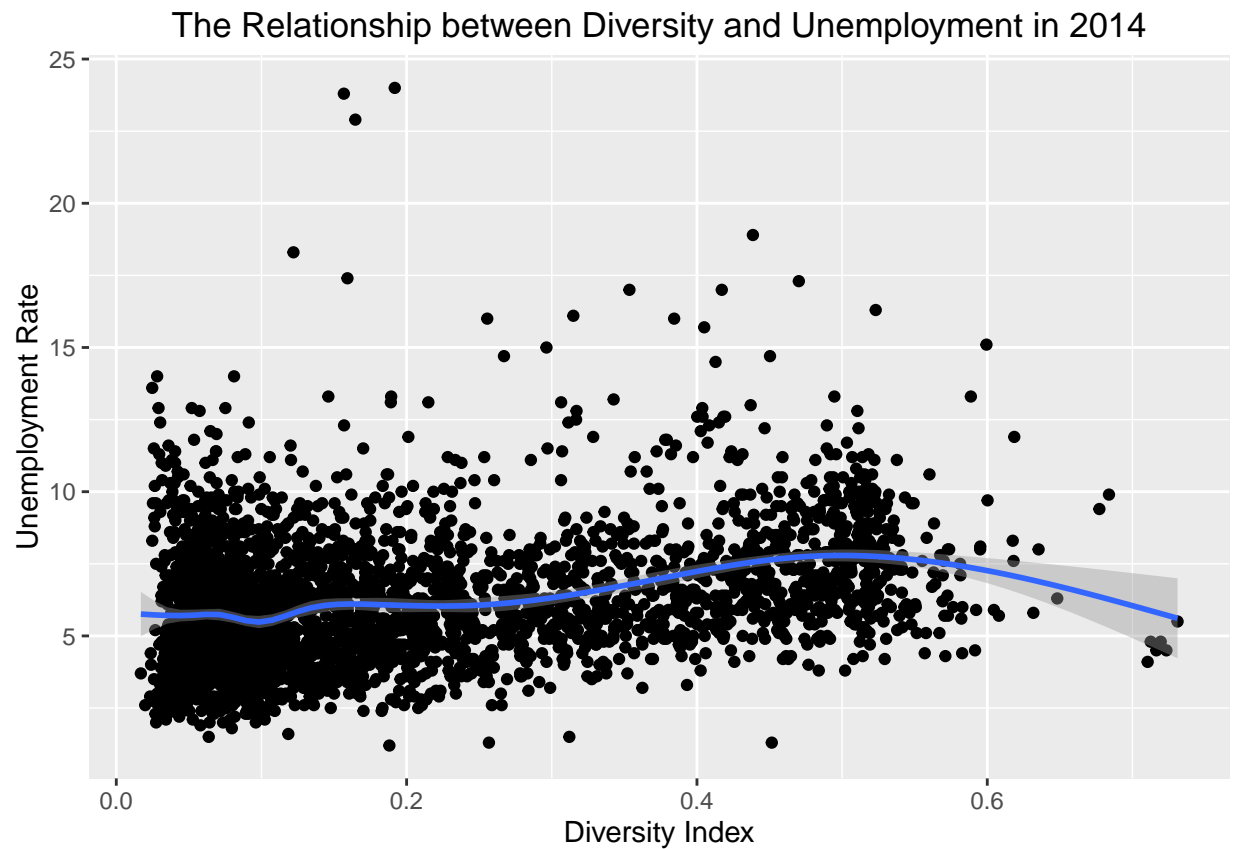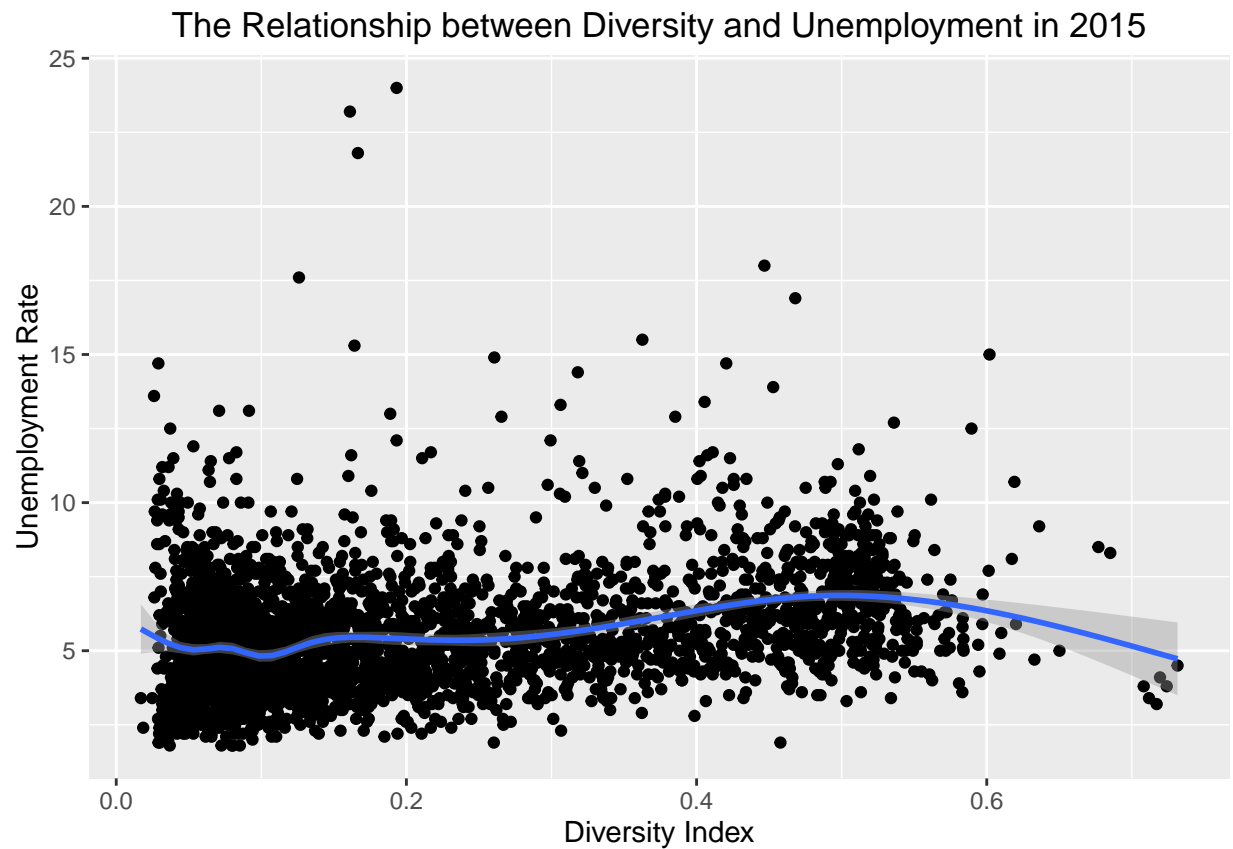
The Relationship between Diversity and Income in 2014


```
##Unemployment Rate
eco_comb2014 %>%
  ggplot(aes(x=INDEX,
             y=Unemployment_rate_2014))+
  geom_point()+geom_smooth()+
  ylab("Unemployment Rate")+
  xlab("Diversity Index")+
  ggtitle("The Relationship between Diversity and Unemployment in 2014")
```

The Relationship between Diversity and Unemployment in 2014


```
eco_comb2015<-left_join(dat2015, eco, by=c("FIPS"="FIPS_Code"))
eco_comb2015 %>%
  ggplot(aes(x=INDEX,
             y=Unemployment_rate_2015))+
  geom_point()+geom_smooth()+
  ylab("Unemployment Rate")+
  xlab("Diversity Index")+
  ggtitle("The Relationship between Diversity and Unemployment in 2015")
```

The Relationship between Diversity and Unemployment in 2015


```r
poverty<-read.csv("PovertyEstimates.csv", header = TRUE, stringsAsFactors = FALSE)
##Poverty
poverty_comb2014<-left_join(dat2014, poverty, by=c("FIPS"="FIPStxt"))
poverty_comb2014 %>%
  ggplot(aes(x=INDEX,
             y=PCTPOVALL_2014))+
  geom_point()+geom_smooth()+
  ylab("Poverty Rate")+
  xlab("Diversity Index")+
  ggtitle("The Relationship between Diversity and Poverty in 2014")
```

The Relationship between Diversity and Poverty in 2014
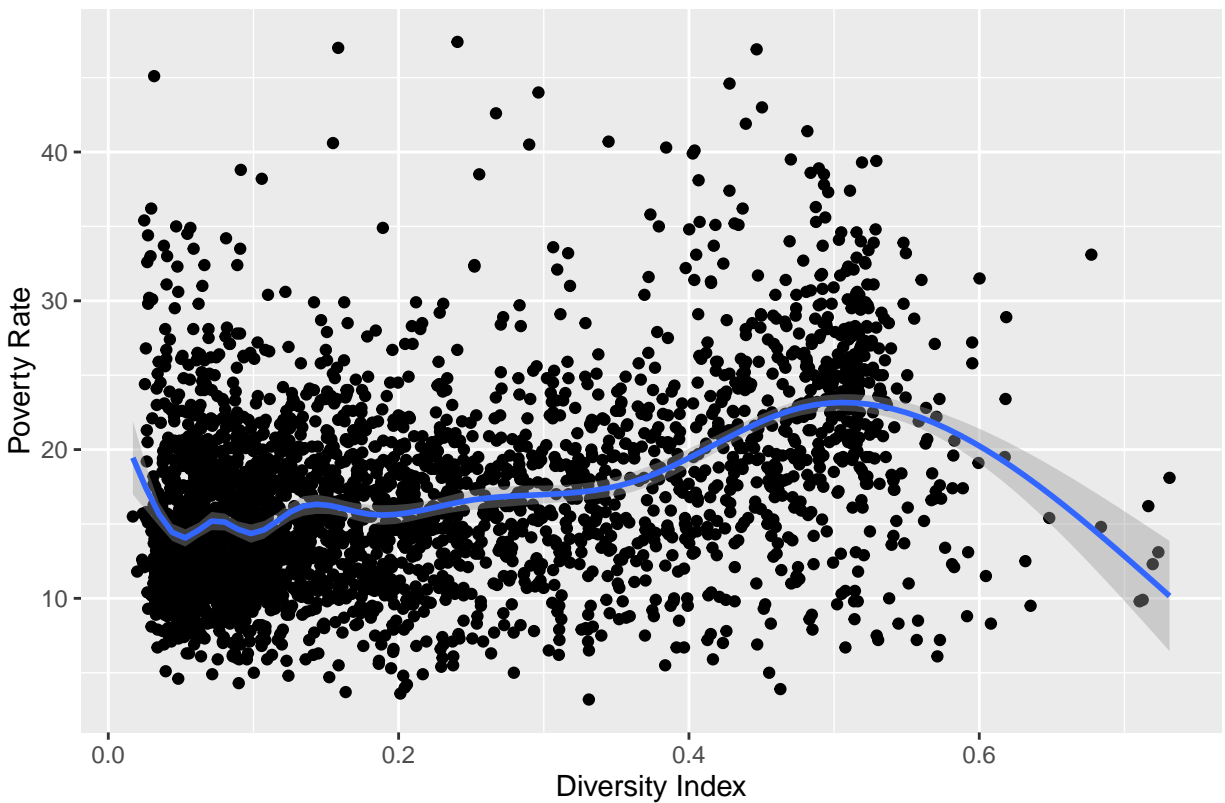

```
##Boxplots

##Generate Diversity Levels
dat2010b<-dat2010 %>%
  mutate(group=0)

for (i in 1:nrow(dat2010b)) {
  if(dat2010b$INDEX[i]>0 & dat2010b$INDEX[i]<0.2)
    dat2010b$group[i]="Level 1"
  else if (dat2010b$INDEX[i]>=0.2 & dat2010b$INDEX[i]<0.4)
    dat2010b$group[i]="Level 2"
  else if (dat2010b$INDEX[i]>=0.4 & dat2010b$INDEX[i]<0.6)
    dat2010b$group[i]="Level 3"
  else
    dat2010b$group[i]="Level 4"
}

dat2011b<-dat2011 %>%
  mutate(group=0)

for (i in 1:nrow(dat2011b)) {
  if(dat2011b$INDEX[i]>0 & dat2011b$INDEX[i]<0.2)
    dat2011b$group[i]="Level 1"
  else if (dat2011b$INDEX[i]>=0.2 & dat2011b$INDEX[i]<0.4)
    dat2011b$group[i]="Level 2"
  else if (dat2011b$INDEX[i]>=0.4 & dat2011b$INDEX[i]<0.6)
```

```
      dat2011b$group[i]="Level 3"
  else
    dat2011b$group[i]="Level 4"
}

dat2012b<-dat2012 %>%
  mutate(group=0)

for (i in 1:nrow(dat2012b)) {
  if(dat2012b$INDEX[i]>0 & dat2012b$INDEX[i]<0.2)
    dat2012b$group[i]="Level 1"
  else if (dat2012b$INDEX[i]>=0.2 & dat2012b$INDEX[i]<0.4)
    dat2012b$group[i]="Level 2"
  else if (dat2012b$INDEX[i]>=0.4 & dat2012b$INDEX[i]<0.6)
    dat2012b$group[i]="Level 3"
  else
    dat2012b$group[i]="Level 4"
}

dat2013b<-dat2013 %>%
  mutate(group=0)

for (i in 1:nrow(dat2013b)) {
  if(dat2013b$INDEX[i]>0 & dat2013b$INDEX[i]<0.2)
    dat2013b$group[i]="Level 1"
  else if (dat2013b$INDEX[i]>=0.2 & dat2013b$INDEX[i]<0.4)
    dat2013b$group[i]="Level 2"
  else if (dat2013b$INDEX[i]>=0.4 & dat2013b$INDEX[i]<0.6)
    dat2013b$group[i]="Level 3"
  else
    dat2013b$group[i]="Level 4"
}

dat2014b<-dat2014 %>%
  mutate(group=0)

  for (i in 1:nrow(dat2014b)) {
    if(dat2014b$INDEX[i]>0 & dat2014b$INDEX[i]<0.2)
      dat2014b$group[i]="Level 1"
    else if (dat2014b$INDEX[i]>=0.2 & dat2014b$INDEX[i]<0.4)
      dat2014b$group[i]="Level 2"
    else if (dat2014b$INDEX[i]>=0.4 & dat2014b$INDEX[i]<0.6)
      dat2014b$group[i]="Level 3"
    else
      dat2014b$group[i]="Level 4"
  }

dat2015b<-dat2015 %>%
  mutate(group=0)

for (i in 1:nrow(dat2015b)) {
  if(dat2015b$INDEX[i]>0 & dat2015b$INDEX[i]<0.2)
    dat2015b$group[i]="Level 1"
```

```
    else if (dat2015b$INDEX[i]>=0.2 & dat2015b$INDEX[i]<0.4)
        dat2015b$group[i]="Level 2"
    else if (dat2015b$INDEX[i]>=0.4 & dat2015b$INDEX[i]<0.6)
        dat2015b$group[i]="Level 3"
    else
        dat2015b$group[i]="Level 4"
}

edu_comb2010b<-left_join(dat2010b, edu, by=c("FIPS"="FIPS.Code"))

edu_comb2010b %>%
    ggplot(aes(x=group,
               y=Percent.of.adults.with.less.than.a.high.school.diploma..2010.2014))+
    geom_point()+geom_smooth()+
    ylab("Percentage")+
    xlab("Diversity Level")+
    ggtitle("No High School Diploma Percentage in Different Diversity Levels")
```
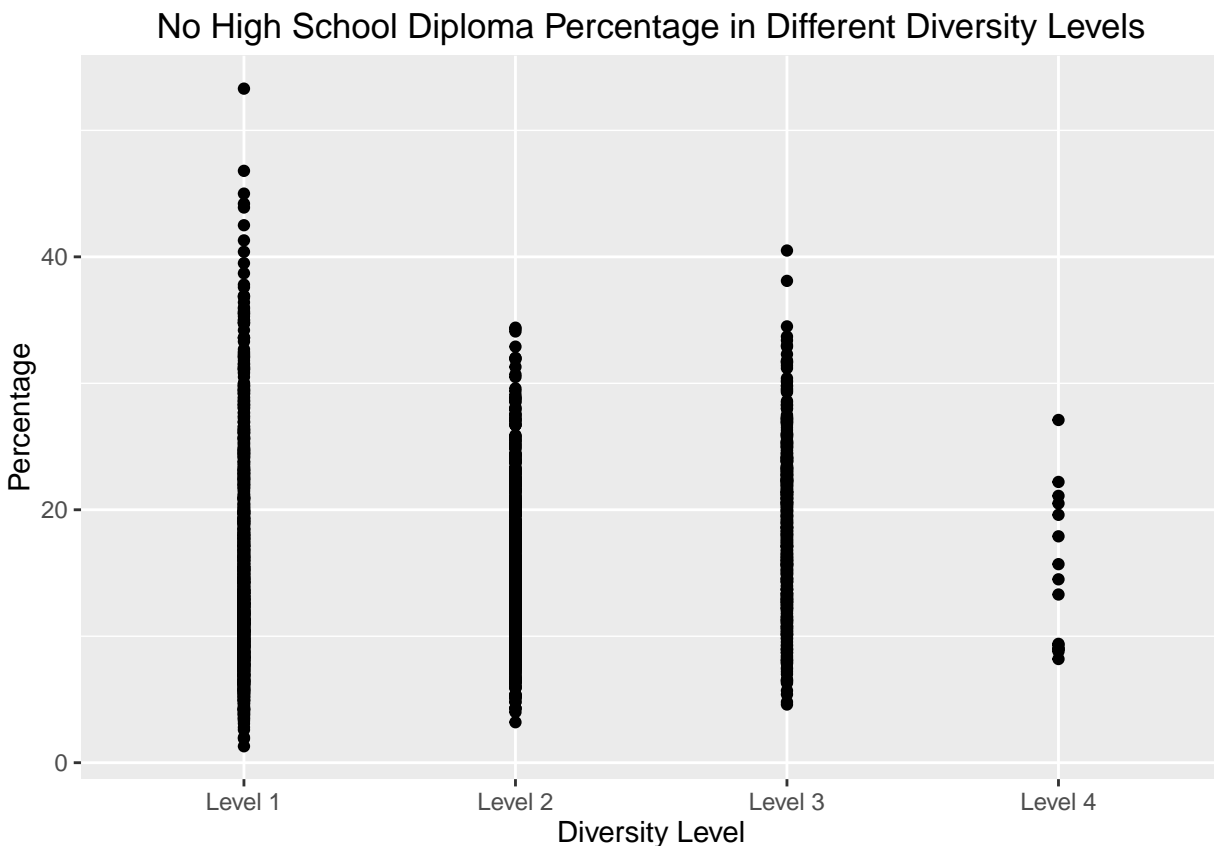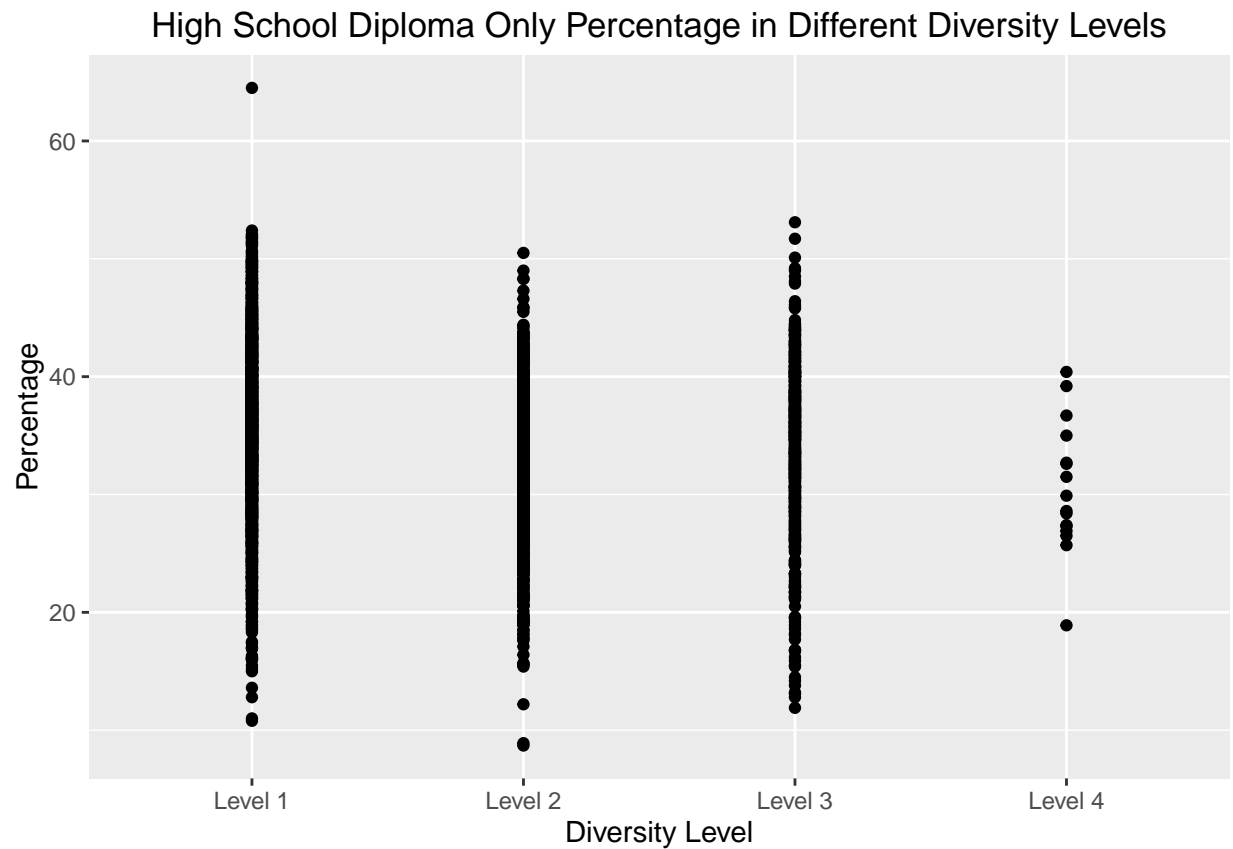

No High School Diploma Percentage in Different Diversity Levels
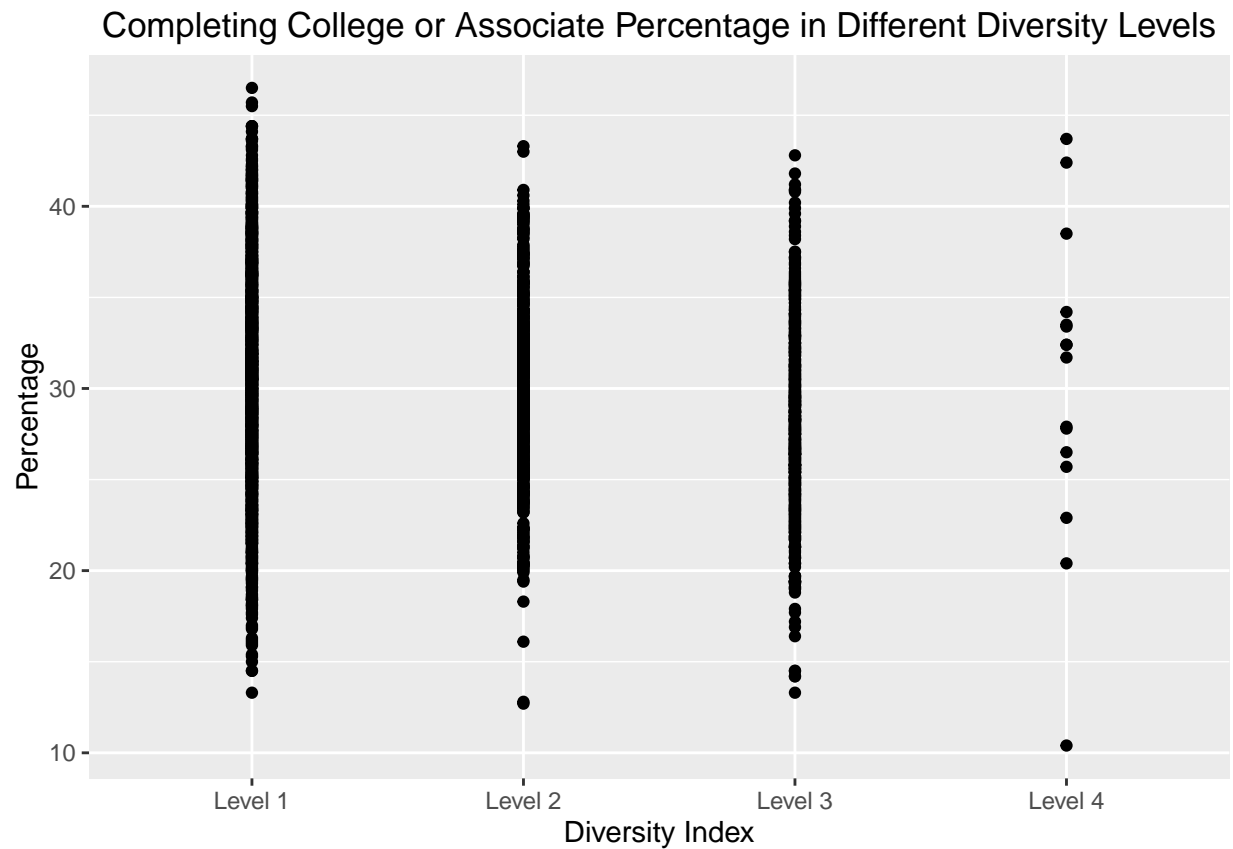
```
edu_comb2010b %>%
    ggplot(aes(x=group,
               y=Percent.of.adults.with.a.high.school.diploma.only..2010.2014))+
    geom_point()+geom_smooth()+
    ylab("Percentage")+
    xlab("Diversity Level")+
    ggtitle("High School Diploma Only Percentage in Different Diversity Levels")
```

High School Diploma Only Percentage in Different Diversity Levels



```
edu_comb2010b %>%
  ggplot(aes(x=group,
             y=Percent.of.adults.completing.some.college.or.associate.s.degree..2010.2014))+
  geom_point()+geom_smooth()+
  ylab("Percentage")+
  xlab("Diversity Index")+
  ggtitle("Completing College or Associate Percentage in Different Diversity Levels")
```

Completing College or Associate Percentage in Different Diversity Levels



```
edu_comb2010b %>%
  ggplot(aes(x=group,
             y=Percent.of.adults.with.a.bachelor.s.degree.or.higher..2010.2014))+
  geom_point()+geom_smooth()+
  ylab("Percentage")+
  xlab("Diversity Index")+
  ggtitle("Bachelor Degree or Higher Percentage in Different Diversity Levels")
```

Bachelor Degree or Higher Percentage in Different Diversity Levels
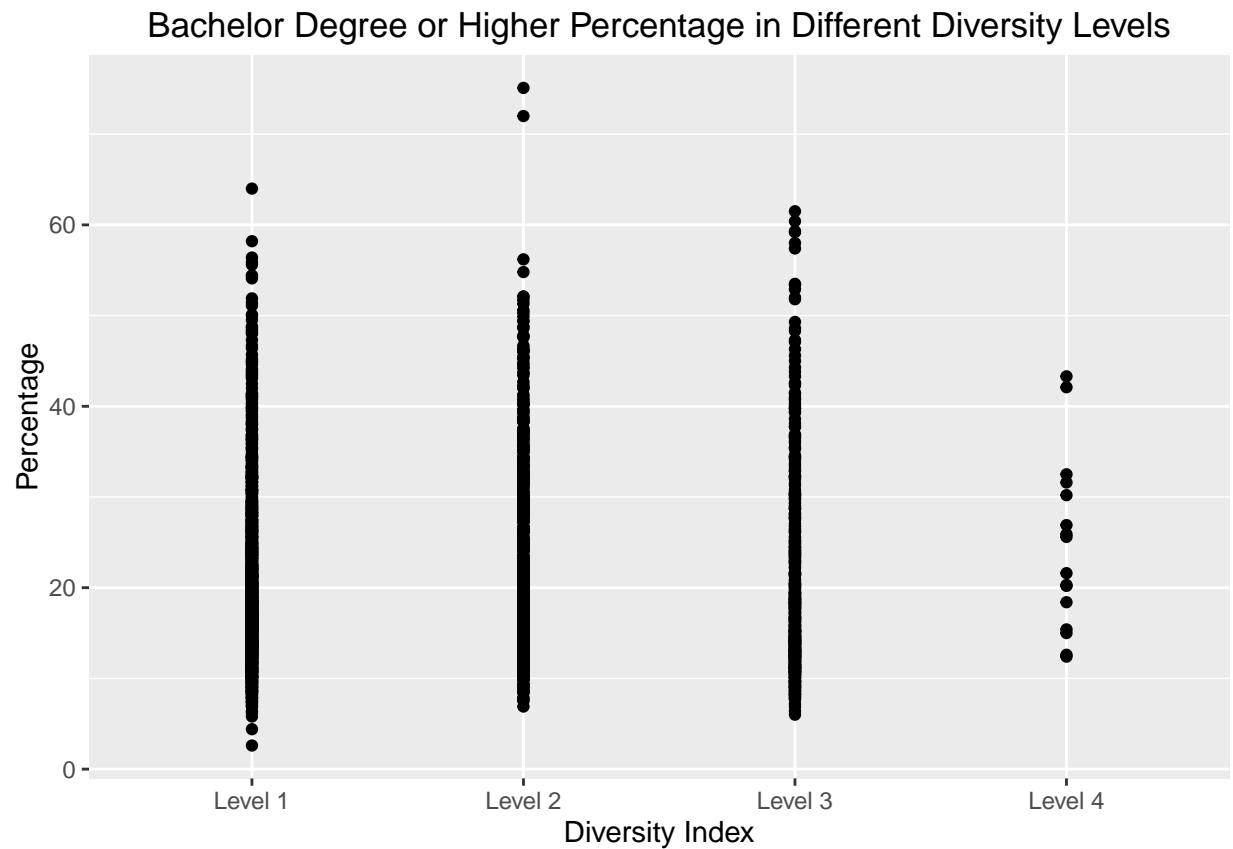

```
eco_comb2014b<-left_join(dat2014b, eco, by=c("FIPS"="FIPS_Code"))
eco_comb2014b %>%
  ggplot(aes(x=group,
             y=Median_Household_Income_2014))+
  geom_boxplot()+
  ylab("Median Income")+
  xlab("Diversity Index")+
  ggtitle("Median Income for Different Diversity Levels in 2014")
```
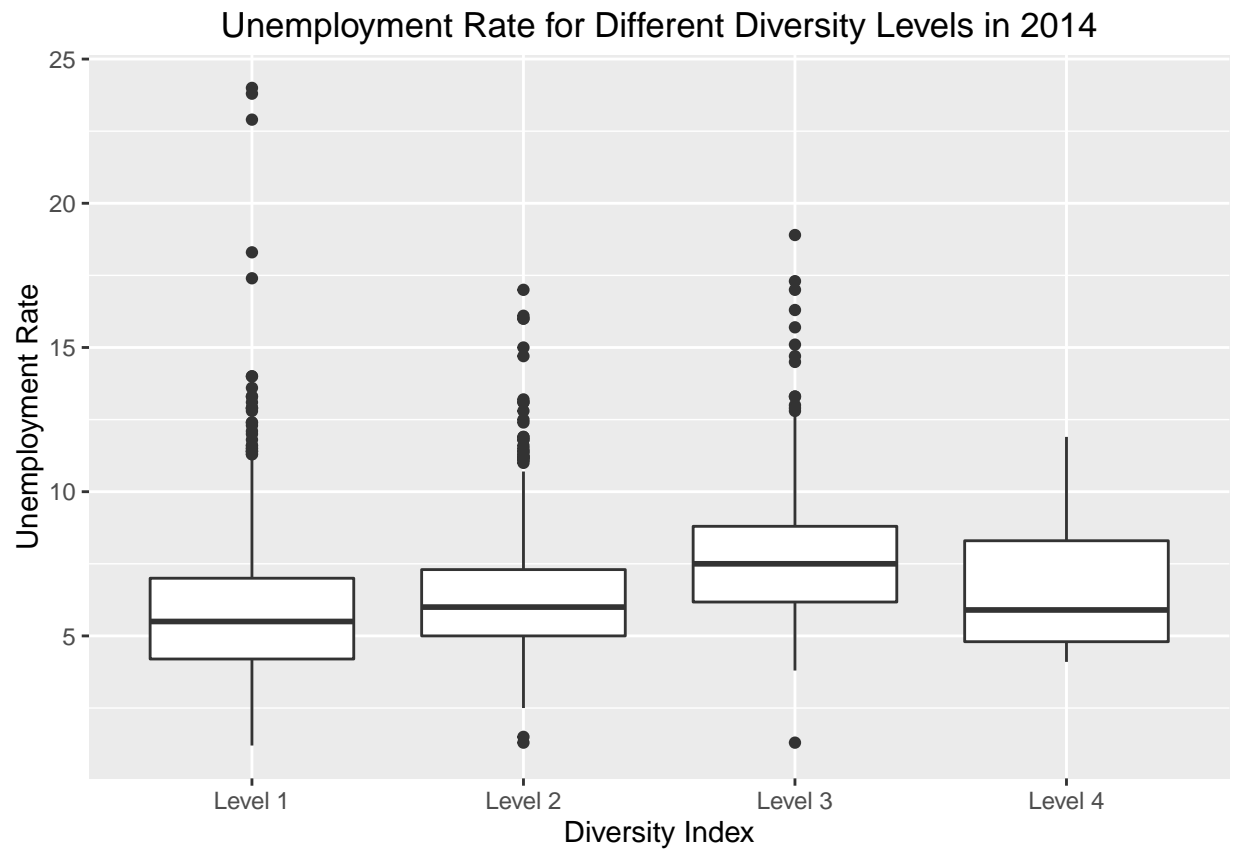
Median Income for Different Diversity Levels in 2014
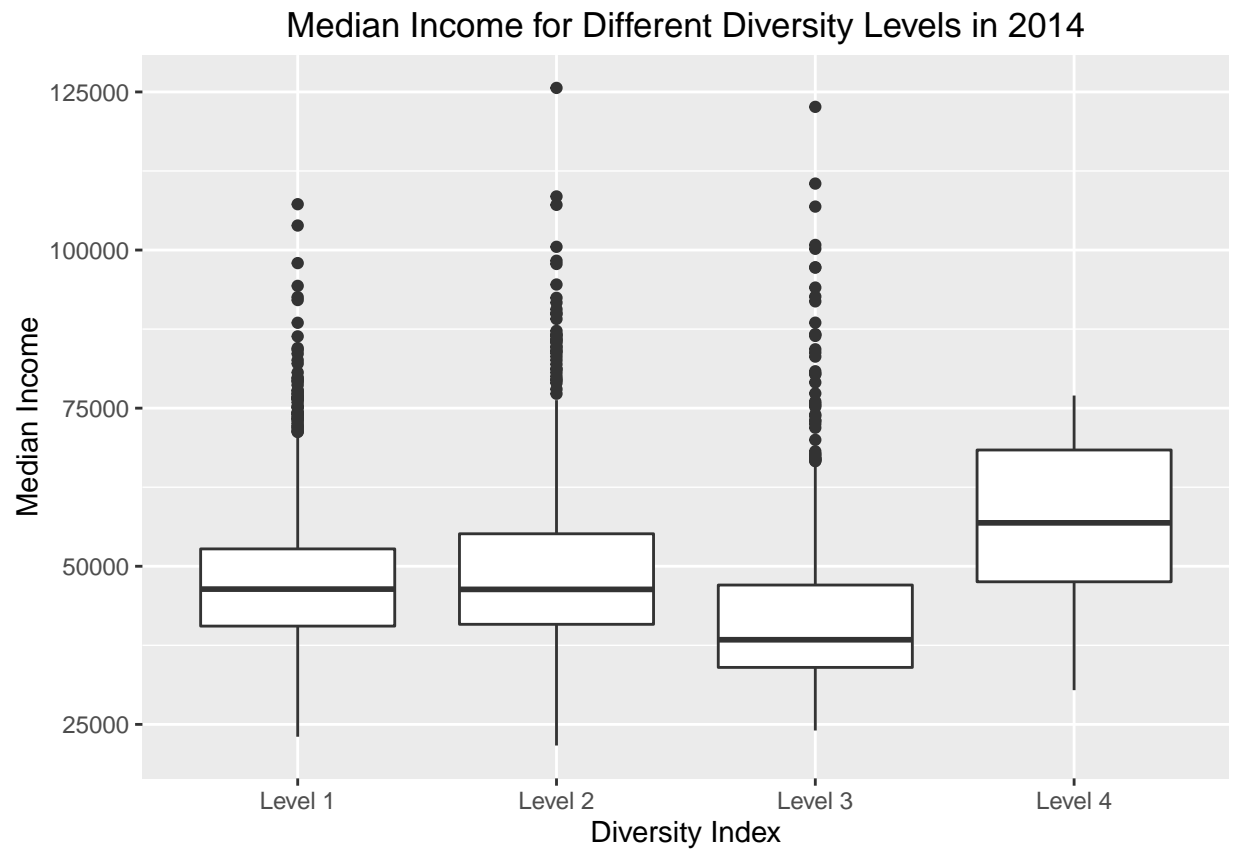


```
eco_comb2014b %>%
  ggplot(aes(x=group,
             y=Unemployment_rate_2014))+
  geom_boxplot()+
  ylab("Unemployment Rate")+
  xlab("Diversity Index")+
  ggtitle("Unemployment Rate for Different Diversity Levels in 2014")
```

Unemployment Rate for Different Diversity Levels in 2014


```
eco_comb2015b<-left_join(dat2015b, eco, by=c("FIPS"="FIPS_Code"))

eco_comb2015b %>%
  ggplot(aes(x=group,
             y=Unemployment_rate_2015))+
  geom_boxplot()+
  ylab("Unemployment Rate")+
  xlab("Diversity Index")+
  ggtitle("Unemployment Rate for Different Diversity Levels in 2015")
```

Unemployment Rate for Different Diversity Levels in 2015
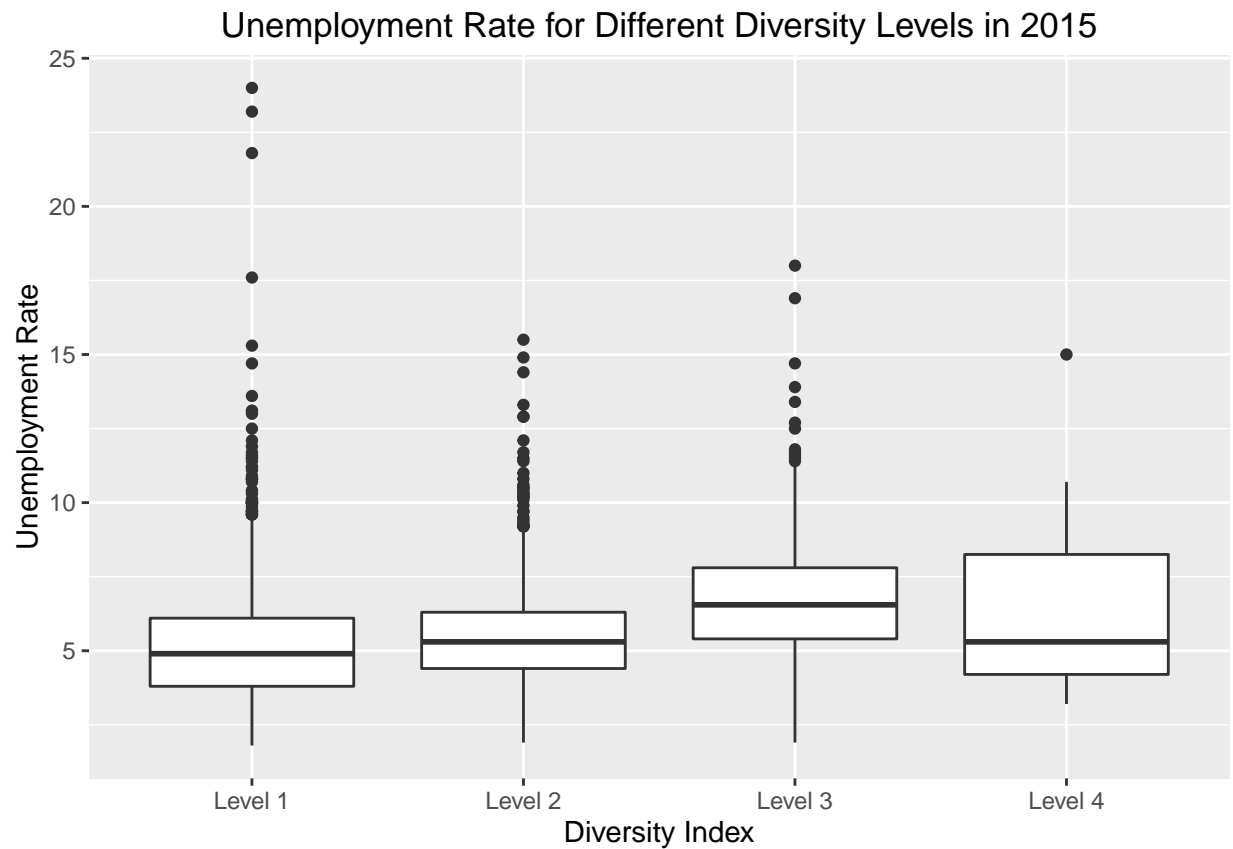
```r
poverty_comb2014b<-left_join(dat2014b, poverty, by=c("FIPS"="FIPStxt"))
poverty_comb2014b %>%
  ggplot(aes(x=group,
             y=PCTPOVALL_2014))+
  geom_boxplot()+
  ylab("Poverty Rate")+
  xlab("Diversity Index")+
  ggtitle("Poverty Rate for Different Diversity Levels in 2014")
```

Poverty Rate for Different Diversity Levels in 2014